DOCUMENT RESUME

ED 121 853 TM 005 287

AUTHOR Yeh, Mary C. L.
TITLE The Issues in the Measurement of Bilingual Language Dominance.
PUB DATE Apr 76
NOTE 16p.; Paper presented at the Annual Meeting of the American Educational Research Association (60th, San Francisco, California, April 19-23, 1976)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage
DESCRIPTORS *Bilingualism; Individual Tests; *Language Proficiency; *Language Tests; *Mexican Americans; *Preschool Children; Preschool Education; Rating Scales; Scoring; Screening Tests; Spanish Speaking; Student Placement; Test Reliability; Test Validity
IDENTIFIERS *Language Dominance

ABSTRACT

This paper deals with measurement of language dominance at the early-childhood level using a rating scale to help bilingual programs with student classification and placement. Some of the assumptions unique in the measurement of language dominance are discussed and applied to the validation procedure on a Spanish/English language dominance scale which is under drastic revision. The instrument was developed for children whose home language may be Spanish, and is designed to be individually administered by a bilingual tester. Part 1 is administered in Spanish and Part 2 in English. The screening discriminates among children in three categories; those preferring English, those preferring Spanish, and bilingual children. The instrument was given to 30 four-year-old Mexican-American children attending three day care centers in the summer of 1974. Teachers' judgments and test results were compared to determine validity, resulting in a coefficient of .86. The test-retest and the rater's reliability coefficients were quite high, which may indicate that the test manual was clearly written, the categories were well defined, and the children performed consistently on this variable. It may be concluded that this is a reliable instrument for four-year-old Mexican-American children at the testing site. (RC)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED121853

THE ISSUES IN THE MEASUREMENT OF BILINGUAL LANGUAGE DOMINANCE

MARY C. L. YEH

SYSTEM DEVELOPMENT CORPORATION

Santa Monica, California

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

paper presented
at

American Educational Research Association

San Francisco, California

April, 1976

TM005 287

Federal and state governments have funded many bilingual programs, the majority of which are designed to provide more efficient instruction to Spanish/English-speaking students. There is a pressing need to develop an objective method for determining student language dominance at the early-childhood or early-elementary level for the purposes of student classification and placement and instructional methods selection. Several measures have been developed for determining language dominance. Zirkel and Greene (1974) listed the four approaches most commonly used to measure bilingualism: (1) reaction time, e.g., Ervin, 1967; (2) word association, e.g., Kolers, 1968; (3) rating scale, e.g., Zirkel, 1971; and (4) flexibility tasks, e.g., MacNamara, 1967.

This paper deals with measurement of language dominance at the early-childhood level using a rating scale to help bilingual programs with student classification and placement. The paper discusses some of the assumptions unique in the measurement of language dominance and applies these assumptions to the validation procedure on a language dominance scale. The assumptions related to bilingual dominance measurement are:

1. A bilingual language dominance test usually involves two languages, in this case, English and Spanish, and each rating scale contains two parts--one in English and one in Spanish. Parallel forms are preferable to direct translation which involves stronger practice effects because learning occurs when the test is administered in the first language. Also, translations may not be equivalent in both languages, and Evans and Guevara (1973) have

shown that direct translation may result in changes of meaning and complexity. Therefore, parallel forms--that is, items of similar difficulty, common to both cultures--should be adopted in order to provide comparable measures of ability and achievement in the bilingual situation (Manuel, 1967).

2. If a test involves two forms, the order of presentation may produce bias because of practice effects. However, Greene and Zirkel (1974) found no significant practice effect in terms of the order of testing when they used the Inter-American Oral Vocabulary subtest as a screening measure of language dominance. They also pointed out that it is essential to follow uniformly one order of administration or the other. More study in this area might be valuable.

3. The judgment of equal difficulty between the two forms of the language dominance test is usually done by the specialist. The difference in test scores between the two parts may not necessarily reflect the difference in difficulty level. The effects of difficulty level always interact with the student's degree of proficiency in either language. The criterion listed in the classical measurement theory (Ghiselli, 1965) for parallel tests is not applicable to the language dominance parallel forms. Further studies in this area are needed.

4. Bilingualism should be thought of as a continuum (Brooks, 1968), even though the purpose of the instrument is to classify the student as "English," "Bilingual," or "Spanish." The three categories are arbitrary, not discrete, and they transition from one stage to another. The line of delineation between the categories is flexible and depends upon the criterion determined under specific circumstances.

5. The language dominance test is not a proficiency measurement, but, rather, measures only the relative degree of preference between the two languages. The individual is compared with himself in terms of his dominance between the two languages regardless of how proficient he is in the languages. The delineation between the classifications depends upon the criterion set up by each program. Many of the traditional measurement principles developed for norm-referenced measurement are not applicable to this kind of instrument; therefore, the validation may not require a norming procedure.

6. The most commonly used measurement approaches (Zirkel and Greene) all involve the so-called "automatic behavior" assumption (Warshaw, 1934, and Dundel, 1948). The stronger language should automatically be used when response is scored by speed and free word association. The use of the dominant language should also be more flexible for very young children who will automatically respond to or with the dominant language more often and faster.

A SAMPLE VALIDATION STUDY

A Spanish/English language dominance scale is being developed to determine preschool children's language preferences for initial instruction. The instrument is under drastic revision and the results of this study do not represent the instrument itself; however, the approach used may serve as an example for the development of a bilingual language dominance test. This instrument was developed for children whose home language may be Spanish, and is designed to be individually administered by a bilingual tester. Part I is administered in Spanish and Part II in English. The screening discriminates among children in three categories: (1) children who prefer English; (2) children who prefer Spanish; and (3) bilingual children.

Scoring System

Each item can be scored as 0 to 4 (i.e., "no response" to "extensive talk"). According to the sixth assumption mentioned above, children will automatically respond to or with the preferred language more often and faster. A positive score is given if a child responds in the administered language. A negative score is given if he responds in the other language (see Figure 1). Each item is scored, and by summarizing across items on one part of the test, a subscore on the language of administration for that part is derived. Therefore, each child receives one score for the Spanish part and one for the English. Since language preference is the primary concern, the difference between the two scores (S-E; Spanish-English) yields a total score for each child.

According to the fourth assumption above, bilingualism should be thought of as a continuum, and the difference score (S-E) distributes along a continuum with 0 in the middle (see Figure 2). Assuming the two forms are of equal difficulty, if S-E=0, this indicates that the individual has about equal preference for both languages. Note that a positive total score indicates stronger dominance in Spanish and a negative total score indicates stronger dominance in English. The greater the difference in scores, the greater is the student's dominance in one language.

Methods

Subjects: The subjects were 30 four-year-old Mexican-American children attending three day care centers in the summer of 1974.

Testers: Six bilingual testers and one bilingual specialist administered the test to children of whom they had no previous knowledge.

Design: The basic design involved a test-retest procedure of the 30 children (see Figure 3). Fourteen children, randomly selected, were initially tested by a linguistic specialist (designated as X) and retested by bilingual testers. Sixteen were tested initially by the bilingual testers and retested by the specialist. At the end of each administration, the testers were asked to judge the child's language dominance based on the test experience. Scoring was done on the computer after all data were collected. The time lapse between test and retest was one to five days. All teachers were asked to judge the language dominance of children in their own classes according to the definition provided in the test manual at the end of the summer semester. The judgment was based on

their intensive daily experience with the children and was to be independent of the test results. All initial test sessions were tape recorded; the tapes recorded by the linguistic specialist were rated by the specialist, and those recorded by other testers were rated by two bilingual raters (designated as Y and Z).

Scoring System Validation: The examiners were asked to judge children's language preferences at the end of each administration independent of the test scores. Therefore, a comparison between the test score and tester's judgment provides a validation of the scoring system. The total scores for this sample of subjects (i.e., S-E) range from +60 to -80. Of the 30 children, only one was judged to prefer Spanish, 10 were judged as bilingual, and 19 were judged as English. (Figure 1 presents this comparison.) If "+40" and "-40" are arbitrarily considered the cut-off points of the three categories, a child who scores "+40", or more, has a stronger dominance in Spanish; a child who scores between "+40" and "-40" is classified as bilingual, and a child who scores "-40" or less has a dominance in English. Among the 30 children tes ted, only one who scored less than "-40" was judged by the tester as bilingual. The remaining 29 children were classified into the three categories consistently by both judgment and test scores; one as Spanish, nine as bilingual, and 19 as English. This indicates that the scoring system is consistent with the testers' judgment.

Parallel Forms

According to the first assumption above, two parallel forms are preferred to two identical forms by translation. The instrument was so constructed. According to the second assumption, the presentation order may effect test scores; so, in this study, the administration follows a uniform order (Spanish first, English

second). To detect the effects of order of presentation, it is recommended that a study such as that by Greene and Zirkel (1974) be duplicated when using this instrument.

According to the third assumption, the items in both forms should be of similar difficulty and common to both cultures. Linguistic specialists selected test items to support this assumption. Some hard data may provide even stronger support. One approach is to have a group of bilingual subjects, whose test scores are around "0," judged in rank order in terms of their Spanish-to-English dominance. The correlation between rank order and test scores may provide a better indication of equivalency between the two forms.

Validity

The purpose of this instrument is to assist the classroom teacher in determining which language will be most effective for teaching the child at the beginning of the school year. If test results can determine a child's language dominance as well as the teacher can, after she has been with the child for a period of time, the test can be considered to measure what it purports to measure. Therefore, test results and teachers' judgments were compared. An analysis of variance was calculated to yield statistical information concerning validity index. The subjects were classified as being Spanish, Bilingual, or English, according to the teacher's judgment. Using each child's total score, a one-way ANOVA was computed on the three groups. (The results are presented in Figure 5.) A ratio between SS_b and SS_t was calculated: $SS_b/SS_t = 0.7365$. The square root of SS_b/SS_t is equal to 0.8582. SS_b/SS_t is the proportion of common variance between

test scores and teacher's judgment to the total variance. The square root of SS_b/SS_t is the validity correlation coefficient (Ghiselli, 1965).

Reliability

Two questions concerning the reliability (Ghiselli, 1965) of the instrument were raised: (1) If a child is administered the same test twice, will the results be the same? (test-retest reliability) and (2) If a child's test performance is rated by different individuals, will they reach the same conclusion? (inter-rater reliability). To answer these questions, the correlation coefficients were calculated on different combinations of test results and independent raters. These are presented in Figure 6, where the first row contains the test-retest reliability figures for the Spanish part, the English part, and the total. They are r=0.945, r=0.850, and r=0.934, respectively. The initial testing session was tape recorded. Fourteen children were tested by the linguistic specialist, and the recording was then rated again by the same specialist after two to three weeks. The reliability coefficients for the two parts and the total were r=0.991, r=0.937, and r=0.987 respectively. This involves only the difference in the same rater at different times. Due to recording problems, only 13 of the 16 initial tests administered by bilingual testers were tape recorded. They were all rated by both Y and Z. The correlation coefficients between the original test results and the results from the recordings rated by Y were: r=0.991, r=0.958, and r=0.986. The correlation coefficients between the original test results and results from the recordings rated by Z were: r=0.987, r=0.970 and r=0.978. The inter-rater reliability coefficients (recordings rated by both Y and Z) were: r=0.993, r=0.967 and r=0.988. These involve consistency between different raters. All the reliability coefficients reported above are very high; therefore, it may be concluded that the instrument is reliable.

Discussion

The design and sample were not ideal because by the time the data were collected, most of the children had been in school for at least six months. Some children who were Spanish speakers or bilinguals at the beginning of school tended to become more proficient in English as the semester progressed, and thus, came to be classified as English speakers. Therefore, the scoring system and the resulting validation data do not represent the real situation under which the test will be used (i.e., the beginning of the school year). Another sample of four-year-old Mexican-American children who have just started school should be tested for a second validation study.

The teacher's judgment was used as the validity criterion, but teachers' judgment is not infallible. The validity coefficient of 0.86 is satisfactory, but if the total sample had included more Spanish and bilingual children, the validity coefficient might be different. As for the reliability of the instrument, the test-retest and the rater's reliability coefficients were quite high, which may indicate that the test manual was clearly written, the categories were well defined, and the children performed consistantly on this variable. It may be concluded that this is a reliable instrument for four-year-old Mexican-American children at the testing site. If it is also intended to use the test nation-wide with five-year-old and six-year-old Mexican-American children, a similar validation study should be carried out for groups of children who have just started school.

According to the fifth assumption above, the language dominance test is not a proficiency measurement. The delineating point for classification will be

determined by the criterion of each individual case. The standard by which an individual's performance is compared is the behavior specified by each program's objective. Therefore, no norming is needed to validate the instrument. The critical measurement issue is how reliably and validly the instrument measures its criterion behavior. The question "Are the two forms parallel?" has not been completely answered in this study as no quantitative data have been collected to support its assumption. Judgment of specialists was used as a criterion for parallelism. The effects of the differences in degree of difficulty and order of presentation between the two forms were investigated, and some approaches were recommended earlier. However, some better approaches are needed.

This sample validation study has illustrated how the six assumptions in validating a bilingual language dominance scale can be applied. Certain measurement problems have been noted that should concern those who assess bilingual language dominance.

REFERENCES

Brooks, Nelson, "The Meaning of Bilingualism Today", Foreign Language Annals, 2(1967), 304.

Dunkel, H.B., Second Language Learning, New York: Ginn and Co., 1948.

Evin, S., "An Issei Learns English", Journal of Social Issues,1967, 23, 78-90.

Evans, Joyce and Guevara, A.E., "Classroom Instruction for Young Spanish Speakers", Exceptional Children, Sept., 1974.

Ghiselli, Edwin E., Theory of Psychological Measurement, McGraw-Hill, 1971.

Greene, John F. and Zirkel, Perry A., "The Use of Parallel Testing of Aural Ability as an Indicator of Bilingual Dominance," Psychology in the Schools, Vol. XV, 51-55, Jan., 1974.

Kolers, P.A., "Bilingualism and Information Processing", Scientific American, 1968,218,78-84.

MacNamara, J., "The Bilingual's Linguistic Performance: A Psychological Overview", Journal of Social Issues, 1967, 23, 58-77.

Manuel, H.T., Technical Report: Test of General Ability and Inter-American Series; Forms CE, DE, CES, DES. Austin, Texas: Guidance Testing Associates, 1967.

Warshaw, J., "Automatic Reactions in Practical Foreign Language Work", Modern Language Journal, 1934, 9, 151-158.

Zirkel, P.A., Puerto Rican Families and Mainland Schools, Hartford, Conn: Hartford Model Cities, 1971.

Zirkel, Perry A. and Greene, John F., "The Validation of Parallel Testing of Aural Ability as an Indicator of Bilingual Dominance", Psychology in the Schools, Vol. XI, 153-157, April, 1974.

Figure 1, Item Scoring System

language of response \ language of administration	Spanish	English
Spanish	0, 1, 2, 3, 4	0,-1,-2,-3,-4
English	0,-1,-2,-3,-4	0, 1, 2, 3, 4

Figure 2, Bilingualism as a Continuum

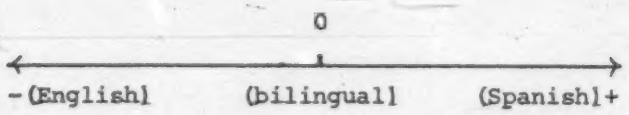

Figure 3, Experimental Design

TEST (recorded)		RETEST	
N=14, tested by a bilingual specialist (X) -the same tester's judgment based on test performance -test score -recording rated by (X)	ONE TO FIVE DAYS	N=14, retested by the bilingual testers -the same tester's judgment based on test performance -test score	END OF THE SEMESTER Teacher's judgment based on daily observation
N=16, tested by bilingual testers -the same tester's judgment based on test performance -test score -recording rated by (Y&Z)		N=16, retested by the bilingual specialist (X) -the same tester's judgment based on test performance -test score	

Figure 4, Scoring System versus Tester's Judgment

Judgment	E=19, B=1	B=9	S=1
Test Results	N=20	N=9	N=1

-80 -40 0 +40 +80
English bilingual Spanish

Figure 5, ANOVA on Total Score

Source of Variation	df	Sum of Squares
Total	29	45079.47
Between Groups	2	33202.46
Remainder	27	439.89
SS_b / SS_t		0.74
Square Root of SS_b / SS_t		0.86

Figure 6, Reliability Coefficients

Combinations	Spanish	English	Total
Test-Retest	r=0.945	r=0.850	r=0.934
Test & Rater X	r=0.991	r=0.937	r=0.987
Test & Rater Y	r=0.991	r=0.958	r=0.986
Test & Rater Z	r=0.987	r=0.970	r=0.987
Inter Rater (Y&Z)	r=0.993	r=0.967	r=0.988